IPGP DGRF 2015 candidate model evaluation for IGRF-13

G. Hulot, P. Vigneron

Abstract

This note reports on our an assessment of the 11 DGRF 2015 candidate models for IGRF-13, using one month of ASM intensity data from the Alpha and Bravo satellites of the Swarm mission, centred on January 1, 2015. We first specify the test data used and explain the rational for using such data. We next explain the way the test is being conducted. We finally discuss the results, concluding that most models do an essentially equally good job, except for four models that significantly stand out, model G (provided by ISTerre), model H (provided by IZMIRAN), model N (provided by the Spanish team) and model O (provided by Strasbourg). We note that three of these models (G, H and O) are the only one sharing the fact that they do not make direct use of (Swarm) satellite data, but rely on so-called "virtual observatories" (or analogous, in the case of model H) derived from such data. This strategy thus seems to lead to DGRF candidate models that fail to fully account for the type of intensity data we selected (quiet time, night orbit leg) at satellite altitude.

1) Test data used: ASM scalar data from Alpha and Bravo satellites

The goal of DGRF 2015 is to provide the best possible description of the Main Field for epoch 2015.0 at the Earth's surface and in the near-Earth environment. Candidate models are thus expected to be able to account for the intensity data provided by the ASM instruments on the Alpha and Bravo satellites of the Swarm mission within a narrow time window centred on 2015.0 (Charlie no longer had a functional ASM in 2015.0), provided enough selection/corrections are first introduced to avoid/remove signals from sources not expected to be modelled by the DGRF model. Although Swarm ASM scalar data were used by most candidate models in one way or another (see section 4 below), these data have the advantage of being arguably the most accurate data available for epoch 2015.0. Biases have indeed been estimated in flight to be well within the specification of 0.3 nT (see Olsen et al., 2015), and the latest investigations of the so-called "dBSun effect" known to affect the Swarm magnetic measurements (see Tøffner-Clausen et al., 2016) showed that this effect is not affecting the ASM scalar data by more than a fraction of nT, when satellites are orbiting in nominal conditions (which was the case all the time for the data used here). These data are therefore the most representative of the "true" field at the altitude they were acquired.

To minimize possible contributions from external fields not accounted for by DGRF candidate models, test data were selected using the following criteria:
- Kp index less than 10
- Night side orbit leg based on Local Time

These data were next decimated by a factor 10 (one data every 10s)

To minimize impact of secular variation, only one month of data centred on January 1, 2015 were used. This selection indeed ensures that enough data could be used to ensure global coverage, while also ensuring that secular variation (SV) effects (not taken into account by DGRF candidate models) will not introduce signals larger than a couple of nT. This was checked by using the parent model of the IPGP DGRF candidate model (model F) in the following way: for each scalar test data to be used, we computed the intensity predicted by the model up to degree 13 including SV and next subtracted the intensity predicted by the IPGP DGRF candidate model up to degree 13 (without SV). The resulting map, useful for latter reference is plotted as Figure 1. Although we did not explicitly check, we are confident that very similar maps could be plotted using any parent model from any of the other candidate models, and that Figure 1 can thus be used for the present purpose of assessing the impact of SV on our evaluation strategy. The corresponding "night time" LT then range between 03:36 and 01:04 (mean 02:06) for Alpha, and between 04:57 and 02:30 (mean (04:28) for Bravo.

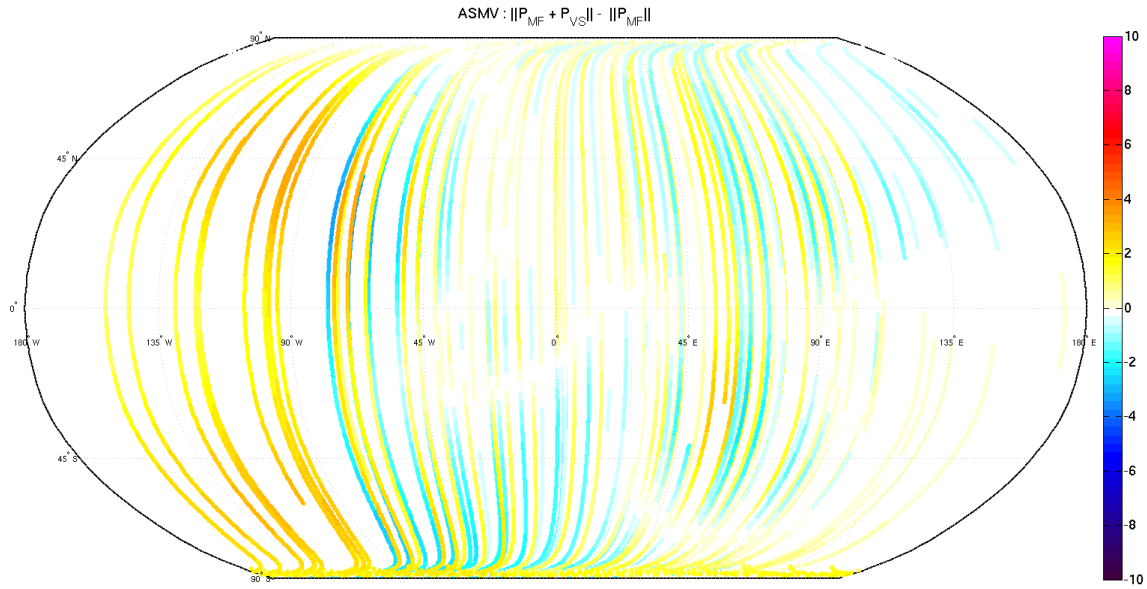

Figure 1: Impact of ignoring SV on the evaluation of DGRF candidate models. Plotted are the residuals between the intensity predicted by the IPGP parent model up to degree 13 with SV and that predicted by the IPGP DGRF candidate model, for all points selected for the evaluation, as described in the text. Colour scales ranging from -10 nT to 10 nT.

2) External field corrections used to account for magnetospheric signals not described by DGRF candidate models

Despite the above data selection, comparing scalar data to DGRF predictions at satellite altitude will mainly reveal magnetospheric signals that are always present, making DGRF candidate model comparison meaningless. To avoid this issue, we introduced magnetospheric field corrections. This was done by relying on the dynamic magnetospheric model co-estimated in the process of building the IPGP DGRF candidate model (see note provided with the IPGP DGRF candidate model for full details). This was done by adding this (vector) prediction to the (vector) prediction of each DGRF candidate model before computing the modulus to be compared to the scalar test data. We are fully aware that this strategy will intrinsically bias our assessment in favour of the IPGP DGRF candidate model, but as we will later see, this external correction appears to be quite successful at removing the unwanted signal from all DGRF candidate models.

3) Crustal Field corrections used to account for crustal signals not described by DGRF candidate models

Crustal field signals also contribute to the intensity data used for our assessment. This is illustrated in Figure 2, where we illustrate the magnitude of this contribution. This map was produced in a way analogous to Figure 1, by using the parent model of the IPGP DGRF candidate model in the following way: for each scalar test data to be used, we computed the intensity predicted by the parent model up to degree 45 (thus adding crustal contribution between degrees 14 and 45, both included) and next subtracted the intensity predicted by the IPGP DGRF candidate model up to degree 13 (without crustal signal). As can be seen, signals of up to roughly 10 nT are found. Contrary to the SV contribution, this becomes significant, and we will later see that such signals indeed tend to mask the signal of interest for assessing DGRF candidate models. We therefore decided to also correct DGRF models for crustal signals. This was done by also adding the (vector) crustal field prediction of the IPGP DGRF candidate model to the (vector) prediction of each DGRF candidate model before computing the modulus to be compared to the scalar test data. This strategy will further bias our assessment in favour of the IPGP DGRF candidate model (we preferred to stick biasing results towards the same model to keep all other models on a reasonably equal foot), but as we will again later see, this crustal correction appears to be quite successful at removing most unwanted signal from all DGRF candidate models.

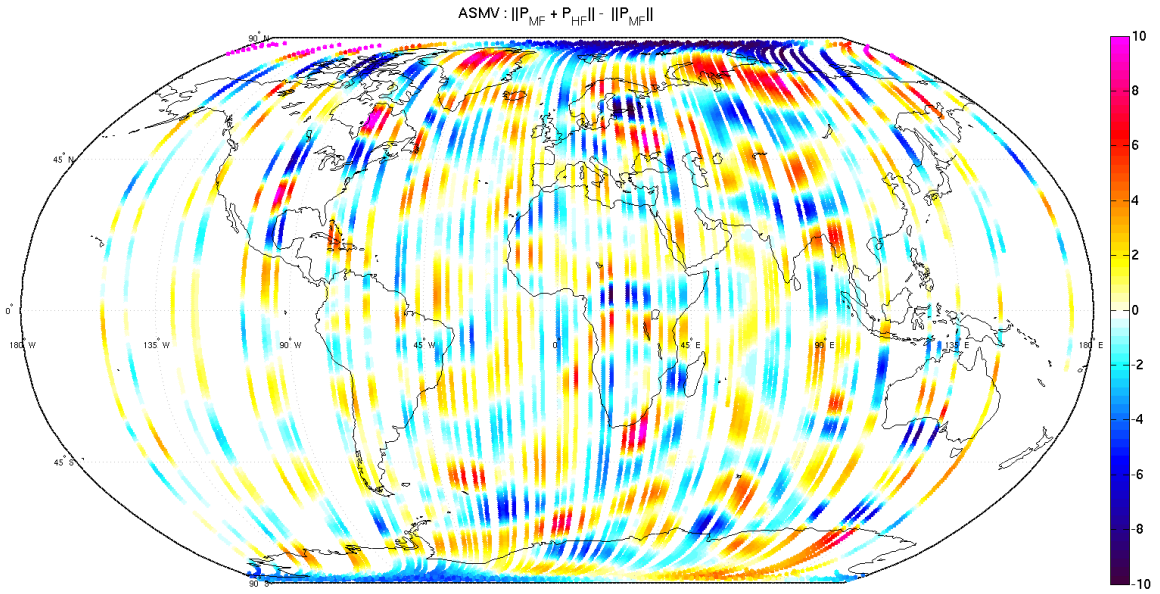

Figure 2: Impact of ignoring crustal field signal on the evaluation of DGRF candidate models. Plotted are the residuals between the intensity predicted by the IPGP parent model up to degree 45 (with crustal field) and that predicted by the IPGP DGRF candidate model, for all points selected for the evaluation, as described in the text. Colour scales ranging from -10 nT to 10 nT.

4) Brief summary of the way Swarm Alpha and Bravo ASM scalar data have been used by the various DGRF candidate models

This summary is inferred from the information provided by the various teams, please refer to the notes they provided, as we may have misunderstood this information.

Note that as a general rule, Swarm Alpha and Bravo VFM vector data having been calibrated with the help of ASM data, their modulus are very close to the values provided by the ASM data.

Model A (BGS) uses Swarm vector data provided by the VFM instruments and observatory ground data. Direct use of ASM scalar data was made only when vector data from the VFM instruments were not available (very few).

Model C (CU/NCEI) uses Swarm VFM vector data up to 55° QD latitude (N and S) and ASM scalar data at all latitudes.

Model D (DTU) uses Swarm VFM vector data and E-W and N-S vector gradient data up to 55° QD latitude (N and S), and ASM scalar and N-S scalar gradient data above these latitudes. The parent model also uses observatory ground data and data from other missions, but only outside of the time-window of interest here, except for Cryosat uncalibrated vector data, used up to December 2014.

Model E (GFZ) only uses Swarm VFM vector data and observatory ground data.

Model F (IPGP) uses ASM experimental vector data from Swarm Alpha and Bravo up to 55° QD latitude (N and S), with occasional use of ASM scalar data, and only ASM scalar data at other latitudes.

Model G (ISTerre) does not make direct use of Swarm data, but relies on virtual observatory data derived from such data (as well as from CHAMP data, but outside the time-window of interest here), together with observatory ground data.

Model H (IZMIRAN) only uses Swarm VFM vector data, but relies on a technique analogous to "virtual observatories", by first computing averages over a grid before model computation.

Model L (NASA/GSFC) uses Swarm VFM vector data and E-W and N-S vector gradient data as well as ASM scalar and N-S scalar gradient data, but in different ways within and outside the +/-55° QD latitudes. The parent model also uses observatory ground data and data from other missions, but only outside of the time-window of interest here.

Model M (Potsdam/MaxPlanck) uses VFM vector data from Swarm Alpha and Bravo. The parent model also uses vector data from the CHAMP mission, but only outside of the time-window of interest here.

Model N (Spanish Team) uses Swarm VFM vector data up to 55° QD latitude (N and S) and ASM scalar data poleward, together with observatory ground data.

Model O (Strasbourg) does not make direct use of Swarm data, but relies on "virtual observatory data" derived from such data, together with observatory ground data.

5) Test results after only taking into account magnetospheric signals not described by DGRF candidate models

Figure 4 shows maps of the differences (residuals) between the intensity test data and the intensity predicted by the various DGRF candidate models after adding contributions from the magnetospheric field as described in section 2. We also computed a mean model, applying equal weight to all models, and show the corresponding residuals for reference.

For all those residuals, we also computed the RMS residual first for all latitudes (Figure 3, left) and next for geographic latitudes between -55° and 55° (Figure 3, right).

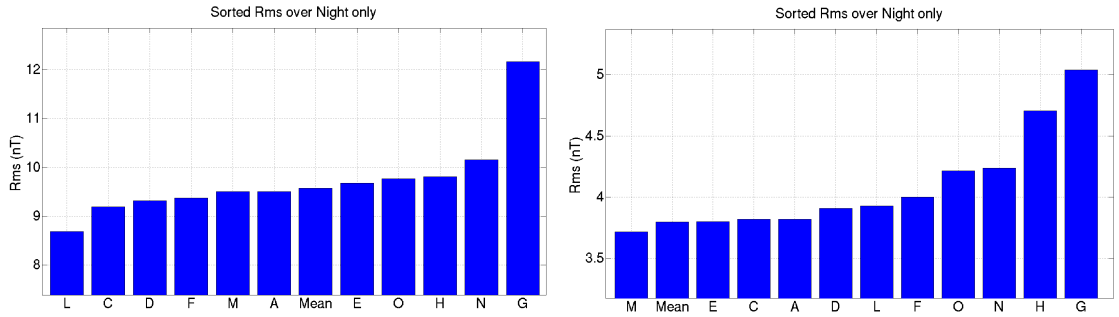

Figure 3: RMS residuals between the intensity test data and the intensity predicted by the various DGRF candidate models (or the mean model) after adding contributions from the magnetospheric field for all latitudes (left) and for latitudes between -55° and 55° (right).

As can be seen, most models lead to very similar residuals, except (from largest to weakest disagreement) models G and N, and to a lesser extent H, O and E, when considering all latitudes, and models G, H, N and O when considering latitudes between -55° and +55°. Also worth noting is that model L appears to distinguish itself as doing significantly better when considering all latitudes, but no longer so (being then comparable to other models) when considering latitudes between -55° and +55°. This ranking is also quite apparent when looking at the maps of Figure 4.

Comparison of these maps that of Figure 2, however, makes it clear that a significant common signal due to the crustal field is present (seen in all residual maps), partly obscuring the way DGRF candidate succeed at accounting for the observations.

6) Test results after also taking into account crustal signals not described by DGRF candidate models

Figure 5 now shows maps of the differences (residuals) between the intensity test data and the intensity predicted by the various DGRF candidate models after adding contributions from both the magnetospheric field (as described in section 2) and crustal field (as described in section 3).

In very much the same way, we again computed the RMS residual first for all latitudes (Figure 6, left) and next for geographic latitudes between -55° and 55° (Figure 6, right).

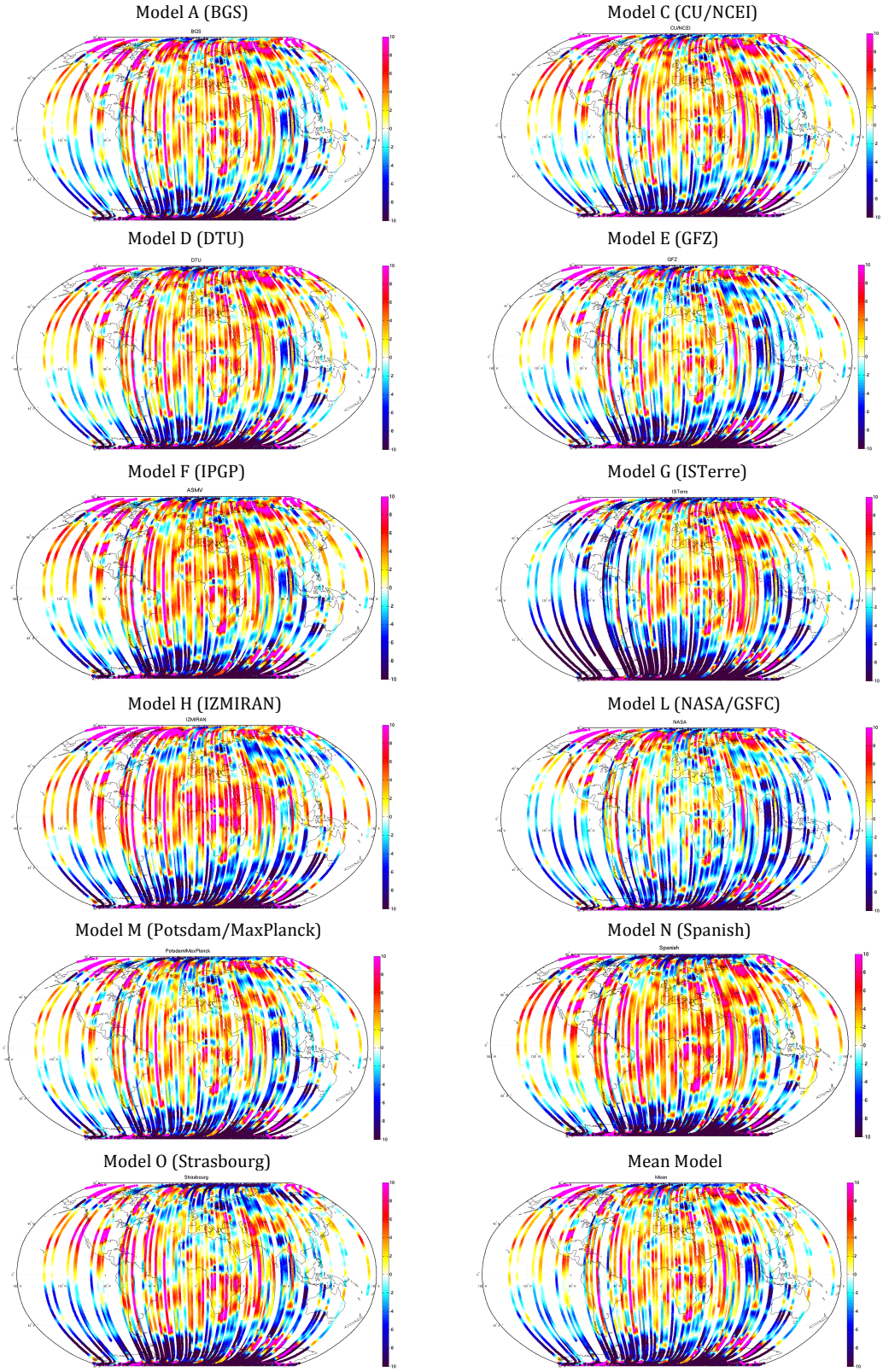

Figure 4: Differences (residuals) between the intensity test data and the intensity predicted by the various DGRF candidate models after adding contributions from the magnetospheric field. Also shown the corresponding map for the mean model. Colour scales ranging from -10 nT to 10 nT.

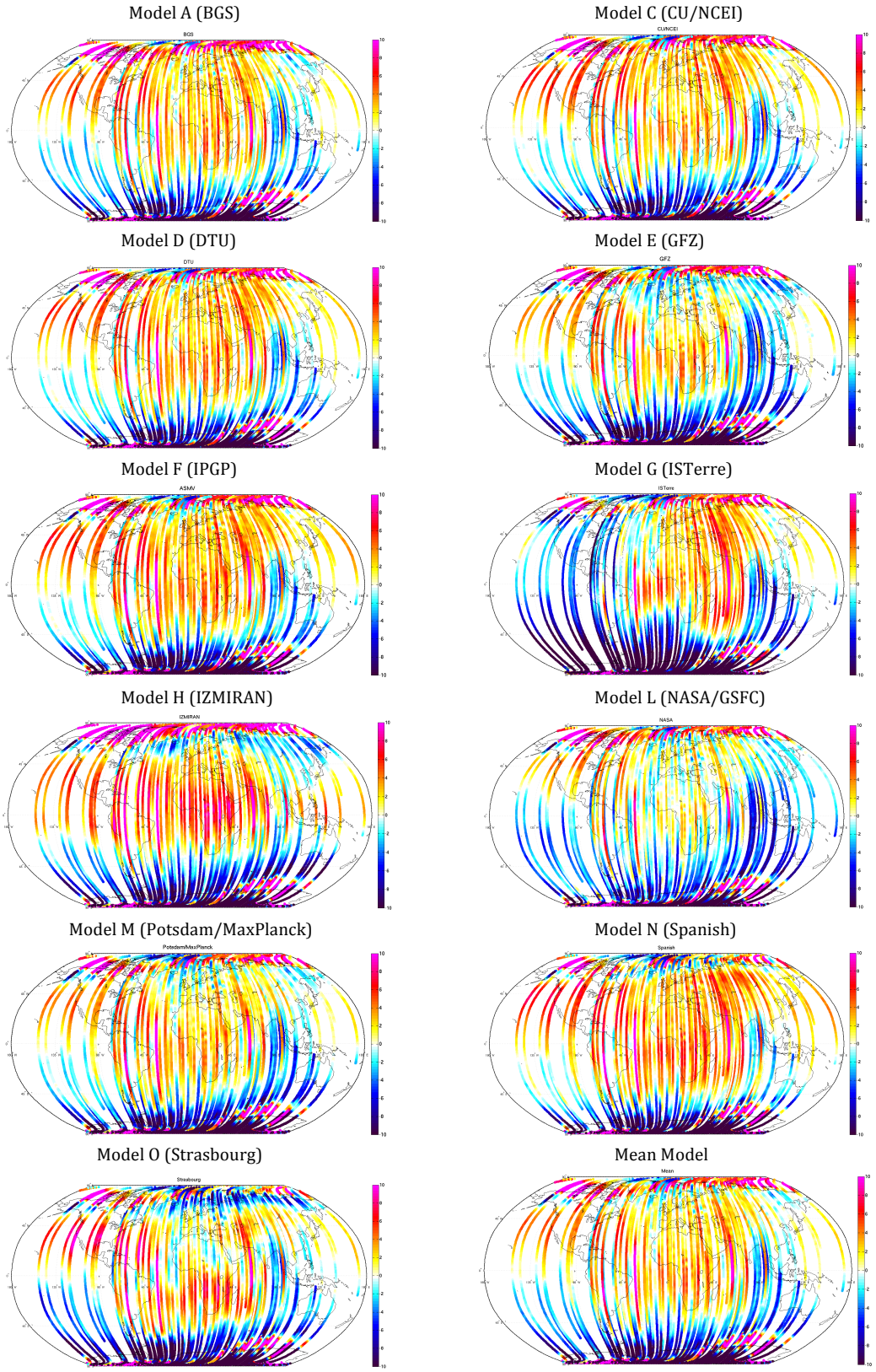

Figure 5: Differences (residuals) between the intensity test data and the intensity predicted by the various DGRF candidate models after adding contributions from both the magnetospheric and crustal fields. Also shown the corresponding map for the mean model. Colour scales ranging from -10 nT to 10 nT.

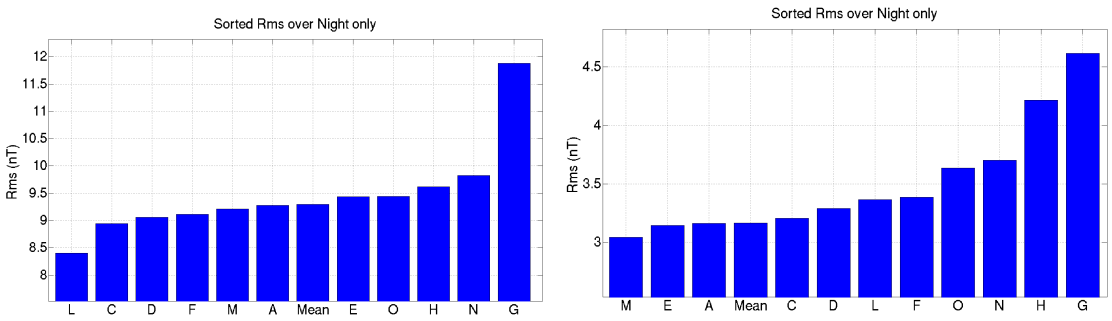

Figure 6: RMS residuals between the intensity test data and the intensity predicted by the various DGRF candidate models (or the mean model) after accounting for contributions from both the magnetospheric and crustal fields for all latitudes (left) and for latitudes between -55° and 55° (right).

As can be seen, and as expected, residuals are now globally reduced. Differences in the way models perform, however, remain very much the same. The ranking is unchanged when considering all latitudes, and only minor changes in the ranking of the best models are observed when considering latitudes between -55° and +55°. We again note that model L is doing significantly better than all other models when considering all latitudes, but not when considering latitudes between -55° and +55°.

This ranking is again also quite apparent when looking at the maps of Figure 5, which now mainly reveals residuals at geographical scales relevant for evaluating the success of DGRF candidate models at accounting for the data. Models with comparable global RMS residuals (both at all latitudes, and at latitudes between -55° and +55°), i.e., models A, C, D, E, F and M, display residuals that share a very similar pattern, very similar also to that of the mean model. Residuals are higher at high latitudes, as one would have expected, because of residual ionospheric signals not taken into account in our test procedure (test data were selected to be on the "night" leg of the orbits, which still includes high latitudes in the Sun, and selection only included simple Kp criteria). These residuals are comparable to what one usually observes with respect to data used when building parent models. Mid-latitude residuals for these models (on order 3nT RMS) are also typical in the same respect. We thus conclude that, from the strict perspective of the present test, models A, C, D, E, F and M can be considered very comparable in terms of their ability to provide reliable DGRF models.

Other models lead to somewhat different geographic distributions of residuals.

Model L does somewhat better at high latitudes than any other of the previous models, and reasonably well at mid-latitude. The geographic pattern is also slightly different, which we interpret as a consequence of the specific modelling strategy used to build the corresponding parent model. Based on the present test, this model too may be considered as reliable as a DGRF candidate model.

Model N does significantly worse at high latitude, and at latitudes between -55° and +55°. The overall behaviour of residuals is similar but somewhat more contrasted than what could be commonly observed in the case of models A, C, D, E, F and M. This we again interpret as a consequence of the specific modelling strategy used to build the corresponding parent model. Based on the present test, this model may be considered as significantly less reliable as a DGRF candidate model.

Model O, which does almost as well as A, C, D, E, F and M at high latitudes, does significantly worse at mid-latitudes, with a different geographic residual pattern. This model is one of the four candidate models that clearly stand out as not doing as well as the other models at mid-latitudes. This we again interpret as a consequence of the specific modelling strategy used to build the corresponding parent model (see more about this below). Based on the present test, this model too may be considered as significantly less reliable as a DGRF candidate model at the critical mid latitudes, where our test may be considered as most robust.

Model H is not doing very well at high latitudes, but clearly stands out as doing significantly worse than all other models (except model G, see below) at mid-latitudes. This is very clearly seen in both Figure 6 and the corresponding residual map of Figure 5, which displays a pattern quite different from that of all other models. This we again interpret as a consequence of the specific modelling strategy used to build the

corresponding parent model (see more about this below). Based on the present test, this model may be considered as one of the least reliable as a DGRF candidate model, particularly at mid latitudes.

Finally, model G clearly stands out as the one doing worse by all standards with respect to the present test. Residuals are much higher than any other models at all latitudes, and particularly at the critical mid-latitudes (doing even worse than models H, N and O, already standing out, as we saw). The modelling strategy of the corresponding parent model is most likely the cause of this discrepancy (see below). From the point of view of the present test, this model may be considered as the single least reliable as a DGRF candidate model.

7) Discussion and Conclusion

Before concluding, it is worth finally plotting the spatial spectra of the differences between the various DGRF candidate models with respect to the average model (with equal weight for all models). This is shown in Figure 7.

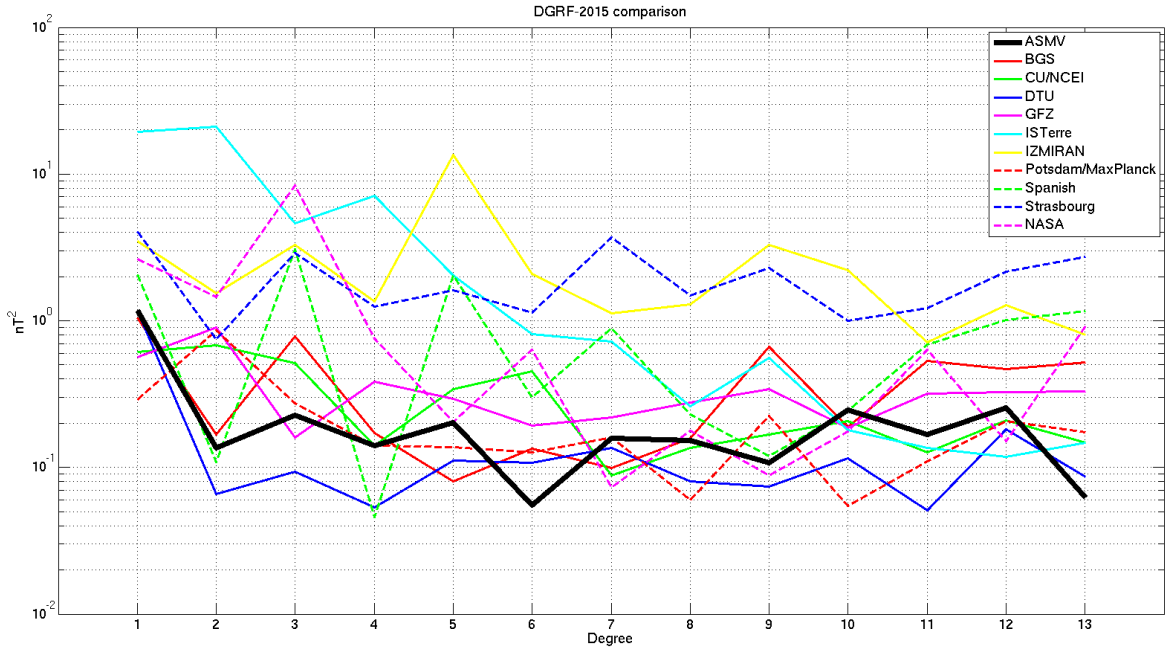

Figure 7: Spatial spectra of the differences between the various DGRF candidate models with respect to the average model (with equal weight for all models), plotted at Earth's surface.

We first note that all models (A from BGS, C from CU-NCEI , D from DTU, E from GFZ, F from IPGP (ASMV), and M from Postdam/MaxPlanck) which passed our test in a similar way (with low RMS values at all and mid-latitudes, see Figure 6) and displayed similar residuals in Figure 5, also display essentially similar spectra.

We next note that model L from NASA/GSFC, which also did very well (particularly so when considering all latitudes), but with different residual patterns (recall Figure 5) slightly stands out for degree 3, which we therefore interpret as being the main cause of the slightly different behaviour of this model with respect to the pattern of residuals. Interestingly, this result also highlights the danger of only relying on such spectral comparisons of models with respect to some average to assess the intrinsic value of a given model. Our test, directly based on observations, does not lead us to consider model L as particularly unreliable, quite the contrary!

Finally, the four models O (from Strasbourg), N (from the Spanish team), H (from IZMIRAN) and particularly G (from ISTerre) identified as most problematic by our test (particularly at the critical mid-latitudes), indeed also display very different spectra. Model G clearly displays particularly large departures at low degrees (up to degree 5), model N at degrees 3, 5 and 7, while both models O and H display systematic strong departures at all degrees.

Based on all these considerations we therefore conclude that out of the 11 DGRF candidate models we tested, 7 appear to be of similarly good value (A, C, D, E, F and M, with the addition of the slightly more different model L), and four appear to be more problematic, particularly at mid-latitudes (O, N, H and G).

It is worth noting that three of the most problematic candidate models from the point of view of the test carried out here are those that did not make direct use of Swarm data. Both models O and G relied on some intermediate data products known as "virtual observatory data" built using Swarm data. Model H relied on a similar concept, using grid averaging of satellite data as their intermediate data product. All other models made direct use of Swarm data. This single difference in the modelling strategy used by the various teams seems to be the main cause of the most critical differences in the way each models reacted to our test.

As the present test was restricted to test the ability of DGRF candidate models to account for intensity data at satellite altitude we therefore conclude that DGRF candidate models derived from strategies based on "virtual observatories" do not seem most adequate for such applications. But our test did not test the value of candidate models for other possible use.

References:

Olsen N., Hulot, G., Lesur, V., Finlay, C.C., Beggan C., Chulliat, A., Sabaka, T.J., Floberghagen, R., Friis-Christensen, E., Haagmans, R., Kotsiaros, S., Lühr, H., Tøffner-Clausen, L., Vigneron, P. The Swarm Initial Field Model for the 2014 geomagnetic field, *Geophys. Res. Lett.*, 42, doi: 10.1002/2014GL062659, 2015.

Tøffner-Clausen, L., Lesur, V., Olsen, N., Finlay C.C., In flight calibration and characterization of the Swarm magnetometry package, *Earth, Planets and Space,* 68:129, doi:10.1186/s40623-016-0501-6, 2016.